AD-A070 209 WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER F/G 12/2
ON THE RATE OF SUPERLINEAR CONVERGENCE OF A CLASS OF VARIABLE M--ETC(U)
APR 79 K RITTER DAAG29-75-C-0024
UNCLASSIFIED MRC-TSR-1950 NL

| OF |

AD A070209

END
DATE FILMED
7-79
DDC

(12) LEVEL II

AD A070209

MRC Technical Summary Report #1950

ON THE RATE OF SUPERLINEAR CONVERGENCE OF A CLASS OF VARIABLE METRIC METHODS

Klaus Ritter

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

April 1979

See 1473 in back

(Received February 26, 1979)

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

79 06 20 037

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

ON THE RATE OF SUPERLINEAR CONVERGENCE OF A CLASS OF VARIABLE METRIC METHODS

Klaus Ritter

Dedicated to Professor Dr. H. Görtler
on the occasion of his seventieth birthday

Technical Summary Report #1950

April 1979

ABSTRACT

This paper considers a class of variable metric methods for unconstrained minimization. Without requiring exact line searches each algorithm in this class converges globally and superlinearly. Various results on the rate of the superlinear convergence are obtained.

AMS (MOS) Subject Classification: 90C30

Key Words: Unconstrained minimization, variable metric method, superlinear convergence.

Work Unit Number 5 - Mathematical Programming and Operations Research.

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

SIGNIFICANCE AND EXPLANATION

Many practical problems in operations research may be reduced to minimizing a function without constraints. In this paper we discuss a class of algorithms for unconstrained minimization problems which converge to the solution from an arbitrary starting point. In order to judge the efficiency of such an algorithm estimates for the rate of convergence are important. Such estimates are derived in this paper.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

ON THE RATE OF SUPERLINEAR CONVERGENCE
OF A CLASS OF VARIABLE METRIC METHODS

Klaus Ritter

1. Introduction

Variable metric methods have been used successfully in unconstrained minimization. Under appropriate assumptions such a method generates a sequence $\{x_j\}$ which converges superlinearly to a global minimum. It is the purpose of this paper to study the rate of the superlinear convergence.

A first result concerning the rate of superlinear convergence of a particular variable metric method, the Davidon-Fletcher-Powell method [4], [6], was obtained by Burmeister [3]. Assuming that the optimal step size is used he proved that this method generates a sequence which converges n-step quadratically when applied to a function $F(x)$ depending on n variables. Using a non-optimal step size Stoer [14] showed that this result is valid for a class of variable metric methods, the so-called restricted Broyden-methods [1], provided the initial point is sufficiently close to a minimizer of $F(x)$. Assuming that for every iteration the last n search directions are uniformly linearly independent and using an appropriate non-optimal step size Schuller [13] proved that the sequence $\{\|x_{j+1}-z\|/\|x_j-z\|\ \|x_{j-n}-z\|$ is bounded, where z is a minimizer of $F(x)$ and the sequence $\{x_j\}$ is generated by the Broyden-Fletcher-Goldfarb-Shanno - method [2], [7], [8], [12].

In this paper we will be concerned with the restricted Broyden methods. These methods have the property that they maintain the symmetry and positive-definiteness of the matrix used to approximate the Hessian matrix of $F(x)$. They form a subclass of the Huang class [9] of variable metric methods. Throughout the paper a non-optimal step size, based on a quadratic interpolation formula, is used.

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

First we will generalize Schuller's result by extending it to all restricted Broyden methods. Secondly we will strengthen Stoer's result by removing the assumption that the initial point has to be close to a minimizer of $F(x)$ and by showing that the sequence $\{\|x_{j+n}-z\|/\|x_j-z\|^2\}$ is not only bounded but converges to zero. Finally assuming that a certain lower bound on the rate of convergence is valid we will show that the sequence $\{\|x_{j+1}-z\|/\|x_j-z\|\ \|x_{j-n+1}\|\}$ is bounded and that the search directions are asymptotically conjugate with respect to the Hessian matrix of $F(x)$ at the global minimizer.

2. A class of variable metric methods

Let $x \in E^n$ and let $F(x)$ be a real-valued function. If $F(x)$ is twice differentiable at a point x_j we denote the gradient and the Hessian matrix of $F(x)$ at x_j by $g_j = \nabla F(x_j)$ and $G_j = G(x_j)$, respectively. A prime is used for the transpose of a vector or a matrix. For any $x \in E^n$, $\|x\|$ denotes the Euclidean norm of x.

We consider the problem of determining a sequence

$$x_{j+1} = x_j - \sigma_j s_j, \quad j = 0,1,2,\ldots \tag{2.1}$$

which converges to a global minimizer z, say, of $F(x)$. Here the vector s_j is called a search direction and the scalar σ_j is referred to as the step size.

If a variable metric method is used to compute the sequence (2.1) then an (n,n)-matrix H_j is associated with each x_j and

$$s_j = H_j g_j \ . \tag{2.2}$$

The matrix H_{j+1} is determined from H_j by adding a rank one or two matrix in such a way that H_{j+1} satisfies the quasi-Newton equation

$$H_{j+1} d_j = p_j \ , \tag{2.3}$$

where

$$d_j = \frac{g_j - g_{j+1}}{\|\sigma_j s_j\|} \ , \quad p_j = \frac{s_j}{\|s_j\|} \ .$$

The various variable metric methods differ in the update procedure which is used to compute H_{j+1} from H_j. A large class of such methods has been studied by Broyden [1], Fletcher [7], Huang [9], and Dixon [5]. In the following we will consider a subclass of these update procedures which ensure that if the initial matrix H_0 is symmetric all matrices H_j are symmetric. It has been shown in

[11] that the update formulas that correspond to this subclass can be written in the form

$$H_{j+1} = H_j + \frac{\beta_1(d_j'p_j + d_j'H_jd_j) + \beta_2 d_j'H_jd_j}{d_j'p_j(\beta_1 d_j'p_j + \beta_2 d_j'H_jd_j)} p_jp_j' \tag{2.4}$$

$$- \beta_1 \frac{p_jd_j'H_j + H_jd_jp_j'}{\beta_1 d_j'p_j + \beta_2 d_j'H_jd_j} - \beta_2 \frac{H_jd_jd_j'H_j}{\beta_1 d_j'p_j + \beta_2 d_j'H_jd_j} \ ,$$

where β_1 and β_2 are arbitrary parameters with $\beta_1^2 + \beta_2^2 > 0$.

Choosing $\beta_1 = 1,\ \beta_2 = 0$ and $\beta_1 = 0,\ \beta_2 = 1$ we obtain the two special cases

$$H_{j+1} = H_j + \frac{d_j'p_j + d_j'H_jd_j}{(d_j'p_j)^2} p_jp_j' - \frac{p_jd_j'H_j + H_jd_jp_j'}{d_j'p_j}$$

and

$$H_{j+1} = H_j + \frac{p_jp_j'}{d_j'p_j} - \frac{H_jd_jd_j'H_j}{d_j'H_jd_j}$$

which are known as BFGS - method (Broyden [2], Fletcher [7], Goldfarb [8], Shanno [12]) and DFP - method (Davidon [4], Fletcher, Powell [6]), respectively.

Assuming that H_0 is positive definite and that, for all j,

$$g_{j+1}'p_j < g_j'p_j, \quad \text{i.e.,} \quad d_j'p_j = \frac{g_j'p_j - g_{j+1}'p_j}{\|\sigma_j s_j\|} > 0$$

we conclude from Lemma 1 in [11] that

$$\beta_1\beta_2 \geq 0 \ , \quad \beta_1 + \beta_2 \neq 0$$

is a sufficient condition for all matrices H_j to be positive definite.

If H_j is positive definite it has been shown in [11] that H_j can be written in the form

$$H_j = \frac{p_jp_j'}{\rho_j g_j'p_j} + \frac{q_jq_j'}{w_j'q_j} + \sum_{i=3}^{n} \frac{p_{ij}p_{ij}'}{d_{ij}'p_{ij}} \ , \tag{2.5}$$

where

i) $p_j = \dfrac{H_j g_j}{\|H_j g_j\|}$, $\rho_j = \dfrac{1}{\|H_j g_j\|}$,

ii) $w_j \in \mathrm{span}\{g_j, g_{j+1}\}$ such that $w_j' p_j = 0$ and $q_j = H_j w_j$ has norm one,

iii) the vectors $d_{3j}, \ldots, d_{nj}$ are orthogonal to p_j and q_j and are such that

$$d_{ij}' H_j d_{kj} = 0, \quad i,k = 3,\ldots,n, \quad i \neq k$$

and

$$p_{ij} = H_j d_{ij}, \quad i = 3,\ldots,n, \text{ has norm one.}$$

Then every H_{j+1} determined by (2.4) has the form (see [11]),

$$H_{j+1} = \frac{p_j p_j'}{d_j' p_j} + \omega_j \frac{u_j u_j'}{w_j' u_j} + \sum_{i=3}^{n} \frac{p_{ij} p_{ij}'}{d_{ij}' p_{ij}} , \tag{2.6}$$

where the vector u_j is uniquely determined by the conditions

$$u_j \in \mathrm{span}\{q_j, p_j\}, \ \|u_j\| = 1, \ d_j' u_j = 0, \ w_j' u_j > 0$$

and the parameter ω_j depends on the particular numbers β_1 and β_2 used in (2.4). More precisely,

$$\omega_j = \gamma_j \left\| q_j - \frac{d_j' q_j}{d_j' p_j} p_j \right\| \tag{2.7}$$

with

$$\gamma_j = \frac{\beta_1 d_j' p_j + \beta_2 \dfrac{(d_j' p_j)^2}{\rho_j g_j' p_j}}{\beta_1 d_j' p_j + \beta_2 d_j' H_j d_j} .$$

We now assume that s_j is determined by (2.2) and H_j is determined by the update formula (2.4) with $\beta_1 + \beta_2 \neq 0$. Under appropriate assumptions on $F(x)$ and on the choice of the step size σ_j it has been shown in [11] that the sequence (2.1) converges globally and superlinearly to a global minimizer of $F(x)$.

In particular if $F(x)$ is convex and twice continuously differentiable and the sequence $\{x_j\}$ converges to some z, say, such that $\nabla F(z) = 0$, $G = G(z)$ is positive definite and the Lipschitz condition

$$\|G(x) - G(z)\| \leq L\|x - z\| \tag{2.8}$$

is satisfied for all x in some neighborhood of z, then it follows from Theorem 3 in [11] that

i) $\dfrac{\|x_{j+1} - z\|}{\|x_j - z\|} \to 0$ as $j \to \infty$.

ii) The sequences $\{H_j\}$ and $\{H_j^{-1}\}$ are bounded,

iii) $\sigma_j = \sigma_j^*$ with

$$\sigma_j^* = \frac{g_j' s_j}{2(F(x_j - s_j) - F(x_j) + g_j' s_j)} \tag{2.9}$$

is an acceptable step size for j sufficiently large, provided $\beta_1 \beta_2 \geq 0$.

It is not difficult to verify that if σ_j^* is defined and positive then it is the global minimizer of the quadratic function $Q_j(\sigma)$ which has the properties

$$Q_j(0) = F(x_j), \ Q_j(1) = F(x_j - s_j), \ \frac{d}{d\sigma} Q(\sigma) = \frac{d}{d\sigma} F(x_j - \sigma s_j) \quad \text{for } \sigma = 0.$$

The above assumptions imply that there are constants $0 < \mu < \eta$ such that for every x in some neighborhood of z,

$$\mu\|y\|^2 \leq y' G(x) y \leq \eta\|y\|^2 \quad \text{for all } y \in E^n . \tag{2.10}$$

Deleting finitely many members of the sequence $\{x_j\}$, if necessary, we may therefore assume that $\sigma_j = \sigma_j^*$ for all j and that there is some neighborhood $U(z)$ of z such that $\{x_j\} \subset U(z)$ and the inequalities (2.8) and (2.10) hold for every $x \in U(z)$.

In view of these results we are justified in requiring that the following assumption is satisfied.

Assumption 1

i) The sequence $\{H_j\}$ is determined by (2.4) with $\beta_1\beta_2 \geq 0$, $\beta_1 + \beta_2 \neq 0$ and H_0 symmetric and positive definite.

ii) The sequences $\{s_j\}$, $\{\sigma_j\}$ and $\{x_j\}$ are determined by (2.2), (2.9), and (2.1), respectively.

iii) There is some z and a neighborhood $U(z)$ such that $F(x)$ is twice continuously differentiable on $U(z)$, $\nabla F(z) = 0$, $\{x_j\} \subset U(z)$ and the inequalities (2.8) and (2.10) are satisfied for every $x \in U(z)$.

iv) $\dfrac{\|x_{j+1}-z\|}{\|x_j-z\|} \to 0$ as $j \to \infty$.

v) There are numbers $0 < \tau_1 < \tau_2$ such that $\tau_1\|x\|^2 \leq x'H_jx \leq \tau_2\|x\|^2$ for all $x \in E^n$ and $j = 0,1,\ldots$.

For later reference we state the following lemma.

Lemma 1

Let Assumption 1 be satisfied. Then the following statements hold.

i) $0 < \gamma_j < 1$ and $\gamma_j \to 1$ as $j \to \infty$.

ii) $|\gamma_j - 1| = O((d_j'q_j)^2) = O\left(\dfrac{\|g_{j+1}\|^2}{\|g_j\|^2}\right)$.

iii) $|g_{j+1}'p_j| = O(\|g_j\|^2)$, $\sigma_j \to 1$ as $j \to \infty$.

iv) $\|d_j - Gp_j\| = O(\|g_j\|)$, $d_j'p_j \geq \mu$, $\|d_j\| \leq \eta$.

v) $\|x_j - z\| = O(\|g_j\|)$, $\|g_j\| = O(\|x_j - z\|)$.

The first three statements of the lemma have been proved in [11]. The last two parts of the lemma are well-known consequences of part iii) of Assumption 1. They can be proved by a simple application of Taylor's theorem.

Remark

It follows from part ii) of Lemma 1 that

$$\frac{\gamma_j - 1}{d_j'q_j} \to 0 \quad \text{as} \quad d_j'q_j \to 0 \ .$$

Therefore, we define $(\gamma_j-1)/d_j'q_j$ to be zero if $d_j'q_j = 0$.

3. Rates of superlinear convergence

Throughout this section we require that Assumption 1 is satisfied.

By (2.3) we have for every j

$$H_{j+1}d_j = p_j,\ d_j = \frac{g_j - g_{j+1}}{\|\sigma_j s_j\|},\ p_j = \frac{s_j}{\|s_j\|} .$$

However, if $i < j-1$ then in general

$$H_j d_i \neq p_i$$

and one of the main difficulties in estimating the rate of superlinear convergence of the sequence $\{x_j\}$ is to find upper bounds for the numbers

$$\|H_j d_i - p_i\|,\ i = j-2,\dots,j-n .$$

As a first step towards this goal we prove the following two lemmas.

Lemma 2

For every j and every $i < j$ let

$$M_j = I - \frac{p_j d_j'}{d_j' p_j} \quad \text{and} \quad v_{ij} = H_j d_i - p_i .$$

Then

i) $$M_j H_j d_j = \frac{1}{\gamma_j}\left[p_j \frac{g_{j+1}' p_j}{g_j' p_j} - p_{j+1} \frac{\|s_{j+1}\|}{g_j' p_j} d_j' p_j\right]$$

ii) $$v_{i,j+1} = M_j v_{ij} + p_j \frac{p_j' d_i - d_j' p_i}{d_j' p_j}$$
$$+ \left(\frac{\gamma_j - 1}{\gamma_j} \frac{d_i' q_j}{d_j' q_j} - \frac{p_j' d_i}{d_j' p_j}\right)\left(p_j \frac{g_{j+1}' p_j}{g_j' p_j} - p_{j+1} \frac{\|s_{j+1}\|}{g_j' p_j} d_j' p_j\right) .$$

Proof:

i) Observing that by (2.5)

(3.1) $$H_j d_j = p_j \frac{d_j' p_j}{\rho_j g_j' p_j} + q_j \frac{d_j' q_j}{w_j' q_j}$$

we obtain

(3.2) $$M_j H_j d_j = p_j \frac{d_j' p_j}{\rho_j g_j' p_j} + q_j \frac{d_j' q_j}{w_j' q_j} - \frac{p_j}{d_j' p_j}\left(\frac{(d_j' p_j)^2}{\rho_j g_j' p_j} + \frac{(d_j' q_j)^2}{w_j' q_j}\right)$$
$$= d_j' q_j \frac{q_j + \alpha_j p_j}{w_j' q_j} \quad , \quad \alpha_j = -\frac{d_j' q_j}{d_j' p_j} .$$

Furthermore, because

$$w_j' p_j = d_j' u_j = 0 \quad , \quad g_{j+1} = g_j - \|\sigma_j s_j\| d_j$$

and by (2.16) in [11]

$$u_j = \frac{q_j + \alpha_j p_j}{\|q_j + \alpha_j p_j\|}$$

it follows from (2.6), (2.7), and (3.2) that

$$s_{j+1} = H_{j+1} g_{j+1} = p_j \frac{p_j' g_{j+1}}{d_j' p_j} + u_j \frac{\omega_j u_j' g_{j+1}}{w_j' u_j}$$
$$= p_j \frac{p_j' g_{j+1}}{d_j' p_j} + (q_j + \alpha_j p_j) \frac{\gamma_j \alpha_j p_j' g_j}{w_j' q_j}$$
$$= p_j \frac{p_j' g_{j+1}}{d_j' p_j} - \gamma_j \frac{p_j' g_j}{d_j' p_j} d_j' q_j \frac{(q_j + \alpha_j p_j)}{w_j' q_j}$$
$$= p_j \frac{p_j' g_{j+1}}{d_j' p_j} - \gamma_j \frac{p_j' g_j}{d_j' p_j} M_j H_j d_j .$$

Therefore,

$$M_j H_j d_j = \frac{1}{\gamma_j}\left[p_j \frac{p_j' g_{j+1}}{p_j' g_j} - p_{j+1} \frac{\|s_{j+1}\|}{p_j' g_j} d_j' p_j\right] .$$

ii) Because by (3.1)

$$\frac{p_j d_j' H_j}{d_j' p_j} = \frac{p_j p_j'}{\rho_j g_j' p_j} - \alpha_j \frac{p_j q_j'}{w_j' q_j}$$

it follows from the definition of H_{j+1} and (3.2) that

$$H_{j+1} = H_j - \frac{p_j p_j'}{p_j q_j' p_j} + \frac{p_j p_j'}{d_j' p_j} - \frac{q_j q_j}{w_j' q_j} + \gamma_j \frac{(q_j+\alpha_j p_j)(q_j+\alpha_j p_j)'}{w_j' q_j}$$

$$= H_j - \frac{p_j d_j' H_j}{d_j' p_j} - \frac{(q_j+\alpha_j p_j) q_j'}{w_j' q_j} + \frac{p_j' p_j'}{d_j' p_j} + \gamma_j \frac{(q_j+\alpha_j p_j)(q_j+\alpha_j p_j)'}{w_j' q_j}$$

$$= M_j H_j + \frac{p_j p_j'}{d_j' p_j} + (\gamma_j - 1) \frac{(q_j+\alpha_j p_j) q_j'}{w_j' q_j} + \gamma_j \frac{(q_j+\alpha_j p_j)\alpha_j p_j'}{w_j' q_j}$$

$$= M_j H_j + \frac{p_j p_j'}{d_j' p_j} + d_j' q_j \frac{(q_j+\alpha_j p_j)}{w_j' q_j} \left[\frac{\gamma_j - 1}{d_j' q_j} q_j' - \frac{\gamma_j}{d_j' p_j} p_j'\right]$$

$$= M_j H_j + \frac{p_j p_j'}{d_j' p_j} + M_j H_j d_j \left(\frac{\gamma_j - 1}{d_j' q_j} q_j - \frac{\gamma_j}{d_j' p_j} p_j\right)' \quad .$$

Hence,

$$v_{i,j+1} = H_{j+1} d_i - p_i$$

$$= M_j v_{ij} + p_j \frac{p_j' d_i - d_j' p_i}{d_j' p_j} + M_j H_j d_j \left(\frac{\gamma_j - 1}{d_j' q_j} q_j' d_i - \frac{\gamma_j}{d_j' p_j} p_j' d_i\right) \quad .$$

In conjunction with part i) this equality completes the proof of the lemma.

Lemma 3

For every $j \geq n$ and $j-n \leq i < \ell \leq j$,

$$v_{i\ell} = 0 \quad \text{if} \quad \ell = i+1$$

$$\left\| v_{i\ell} - \eta_{i\ell} \frac{\|s_\ell\|}{g_{\ell-1}' p_{\ell-1}} p_\ell \right\| = O(\|g_i\|) \quad \text{if} \quad \ell > i+1 \quad ,$$

where

$$\eta_{i\ell} = \tau_{\ell-1} d_i' q_{\ell-1} d_{\ell-1}' p_{\ell-1} - p_{\ell-1}' d_i, \quad \tau_{\ell-1} = \frac{\gamma_{\ell-1}-1}{\gamma_{\ell-1}} \frac{1}{d_{\ell-1}' q_{\ell-1}}$$

and

$$|\eta_{i\ell}| = O(1) \quad , \quad |\tau_{\ell-1}| = O\left(\frac{\|g_\ell\|}{\|g_{\ell-1}\|}\right) \quad .$$

Proof:

Since every H_ℓ satisfies the quasi-Newton equation (2.3) we have $v_{\ell-1,\ell} = H_\ell d_{\ell-1} - p_{\ell-1} = 0$. Let $i < k < \ell$. By part ii) of Lemma 2

(3.3) $$v_{i,k+1} = M_k v_{ik} - \eta_{i,k+1} \frac{s_{k+1}}{g_k' p_k} p_{k+1} + \mu_{i,k+1} p_k \quad ,$$

where

$$\eta_{i,k+1} = \tau_k d_i' q_k d_k' p_k - p_k' d_i \quad , \quad \tau_k = \frac{\gamma_k - 1}{\gamma_k} \frac{1}{d_k' q_k}$$

and

$$\mu_{i,k+1} = \frac{p_k' d_i - d_k' p_i}{d_k' p_k} + \left(\tau_k d_i' q_k - \frac{p_k' d_i}{d_k' p_k}\right) \frac{g_{k+1}' p_k}{g_k' p_k} \quad .$$

Since by Lemma 1, $\|d_j\| \leq \eta$, $\gamma_j \to 1$ as $j \to \infty$,

$$|1-\gamma_j| = O(|d_j' q_j|^2) \quad \text{and} \quad |d_j' q_j| = O\left(\frac{\|g_{j+1}\|}{\|g_j\|}\right)$$

we have

(3.4) $$|\tau_k| = O\left(\frac{\|g_{k+1}\|}{\|g_k\|}\right) \quad \text{and} \quad |\eta_{i,k+1}| = O(1) \quad .$$

Furthermore by Lemma 1, $\|d_j - Gp_j\| = O(\|g_j\|)$ which implies

(3.5) $$|p_k' d_i - d_k' p_i| = O(\|g_i\|) \quad .$$

Therefore,

(3.6) $$|\mu_{i,k+1}| = O(\|g_i\|)$$

because it follows from Lemma 1 that

$$d_k'p_k \geq \mu > 0 \quad \text{and} \quad |g_{k+1}'p_k|/g_k'p_k = O(\|g_i\|) \ .$$

Since $v_{i,i+1} = 0$, the sequence $\{M_j\}$ is bounded and, for every j, $M_j p_j = 0$, the statement of the lemma follows now from (3.4), (3.6) and the equality (3.3).

In order to obtain the first result on the rate of superlinear convergence from the above lemma we make the assumption that, for j sufficiently large, n consecutive search directions are uniformly linearly independent.

Assumption 2

For j sufficiently large P_j^{-1} exists and $\{P_j^{-1}\}$ is bounded, where

$$P_j = (p_{j-1}, p_{j-2}, \ldots, p_{j-n}) \ .$$

Using this assumption and Lemma 3 we can now prove the following theorem.

Theorem 1

If Assumption 2 is satisfied then, for all update formulas (2.4) with $\beta_1\beta_2 \geq 0$, $\beta_1 + \beta_2 \neq 0$, we have

i) $\dfrac{\|x_{j+1}-z\|}{\|x_j-z\|} = O(\|x_{j-n}-z\|)$

ii) $\|H_j - G^{-1}\| = O(\|x_{j-n-1}-z\|)$.

Proof:

i) For j sufficiently large let

$$D_j = (d_{j-1}, d_{j-2}, \ldots, d_{j-n}) \quad \text{and} \quad p_j = P_j y_j \ .$$

Then

$$d_j = GP_j y_j + (d_j - Gp_j)$$

$$= D_j y_j + (GP_j - D_j) y_j + (d_j - Gp_j)$$

and

$$M_j d_j = P_j y_j + (M_j D_j - P_j) y_j + M_j(GP_j - D_j) y_j + M_j(d_j - Gp_j)$$

$$= p_j + (v_{j-1,j}, v_{j-2,j}, \ldots, v_{j-n,j}) y_j + \hat{y}_j \ .$$

where

(3.7) $$\|\hat{y}_j\| = O(\|H_j\|(\|GP_j - D_j\| \ \|P_j^{-1}\| + \|d_j - Gp_j\|))$$

$$= O(\|g_{j-n}\|) \ .$$

Therefore,

$$M_j H_j d_j = M_j (v_{j-1,j}, v_{j-2,j}, \ldots, v_{j-n,j}) P_j^{-1} p_j + M_j \hat{y}_j \ .$$

By Lemma 3 and (3.7) this equality implies

(3.8) $$\|M_j H_j d_j\| = O(\|g_{j-n}\|) \ .$$

Using (3.8) and part i) of Lemma 2 we obtain

$$\frac{\|s_{j+1}\|}{g_j'p_j} \leq \frac{\gamma_j}{d_j'p_j} \|M_j H_j d_j\| + \frac{1}{d_j'p_j} \frac{|g_{j+1}'p_j|}{d_j'p_j} = O(\|g_{j-n}\|)$$

which by Lemma 1 implies

$$\frac{\|x_{j+1}-z\|}{\|x_j-z\|} = O(\|x_{j-n}-z\|) \ .$$

ii) It follows from Lemmas 1 and 3 and the first part of the theorem that, for $i = j-2, \ldots, j-n$,

$$\|v_{ij}\| = O(\|g_{j-1-n}\|) \ .$$

Setting

$$V_j = (v_{j-1,j}, v_{j-2,j}, \ldots, v_{j-n,j})$$

we have

$$(H_j G - I) P_j = V_j - H_j(D_j - GP_j)$$

which implies

$$H_j - G^{-1} = V_j P_j^{-1} G^{-1} - H_j (D_j - GP_j) P_j^{-1} G^{-1} .$$

Therefore,

$$\|H_j - G^{-1}\| = O(\|V_j\| + \|D_j - GP_j\|)$$

$$= O(\|g_{j-n-1}\|) ,$$

which by part v) of *Lemma* 1 completes the *proof of the theorem.*

For the special case $\beta_1 = 1$, $\beta_2 = 0$, i.e., for the Broyden-Fletcher-Goldfarb-Shanno-method the above result has been obtain by Schuller [13].

For the following results we need a recurrence relation for the v_{ij}'s. This will be derived in the following lemma.

Lemma 4

For every $j \geq n$ and $j-n \leq i < k < j$,

(3.9) $$\|v_{i,k+1}\| = O\left(\frac{\|g_{k+1}\|}{\|g_k\|}\left(\frac{|g_k' p_i|}{\|g_k\|} + \|v_{ik}\|\right) + \|g_i\|\right) .$$

Proof

By Taylor's theorem we have

(3.10) $$g_{k+1} = g_k - \sigma_k G s_k - \sigma_k E_k s_k ;$$

where

$$E_k = \int_0^1 G(x_k - t(\sigma_k s_k))dt - G$$

and

(3.11) $$\|E_k\| \leq \max_{0\leq t\leq 1} \|G(x_k - t(\sigma_k s_k)) - G\|$$

$$\leq \max_{0\leq t\leq 1} \|L(x_k - t(x_k - x_{k+1})) - z\|$$

$$\leq L \max\{\|x_k - z\|, \|x_{k+1} - z\|\} = O(\|g_k\|) ,$$

Where the last relation follows from Lemma 1.

Multiplying (3.10) with p_i we obtain

(3.12) $$g_{k+1}' p_i = g_k' p_i - \sigma_k p_i' G s_k - \sigma_k p_i' E_k s_k$$

$$= g_k' p_i - \sigma_k d_i' H_k g_k - \sigma_k (p_i' G - d_i') s_k - \sigma_k p_i' E_k s_k$$

$$= (1-\sigma_k) g_k' p_i - \sigma_k v_{ik}' g_k - y_{ik} ,$$

where

$$y_{ik} = \sigma_k (p_i' G - d_i') s_k + \sigma_k p_i' E_k s_k$$

and because of (3.11) and Lemma 1

(3.13) $$\|y_{ik}\| = O(\|g_k\| \|g_i\| + \|g_k\|^2) = O(\|g_k\| \|g_i\|) .$$

Observing that by definition

$$p_i' d_k = \frac{p_i' g_k - p_i' g_{k+1}}{\|\sigma_k s_k\|}$$

we conclude from (3.12) and (3.13) that

(3.14) $$|p_i' d_k| = O\left(\frac{|g_k' p_i|}{\|s_k\|} + \|v_{ik}\| \frac{\|g_k\|}{\|s_k\|} + \frac{\|y_{ik}\|}{\|s_k\|}\right)$$

$$= O\left(\frac{|g_k' p_i|}{\|g_k\|} + \|v_{ik}\| + \|g_i\|\right) .$$

Next assume that $\|g_{k+1}\| \leq \|g_k\| \|g_i\|$. Then it follows from Lemma 3 that

(3.15) $$|\tau_k d_i' g_k d_k' p_k| \leq n^2 |\tau_k| = O\left(\frac{\|g_{k+1}\|}{\|g_k\|}\right) = O(\|g_i\|) .$$

Now let

(3.16) $$\|g_{k+1}\| > \|g_k\| \|g_i\|$$

and observe that by definition

$$d_i'q_k = d_i'H_k w_k = w_k'p_i + w_k'v_{ik} \tag{3.17}$$

where

$$w_k = \frac{g_{k+1}-\lambda_k g_k}{\|H_k(g_{k+1}-\lambda_k g_k)\|}\ , \quad \lambda_k = \frac{g_{k+1}'p_k}{g_k'p_k}\ . \tag{3.18}$$

Since $w_k'p_k = 0$ it follows from Lemma 3 that

$$|w_k'v_{ik}| = O(\|g_i\|)\ . \tag{3.19}$$

Using (3.16), (3.18) and Lemma 1 we obtain

$$\frac{\|\lambda_k g_k\|}{\|g_{k+1}\|} = \frac{|g_{k+1}'p_k|}{\|g_{k+1}\|}\,\frac{\|g_k\|}{g_k'p_k} = O\left(\frac{\|g_k\|^2}{\|g_{k+1}\|}\right) = O\left(\frac{\|g_k\|}{\|g_i\|}\right)\ .$$

Because $\|g_k\|/\|g_i\| \to 0$ as $j \to \infty$ this relation implies that

$$\left\|H_k\left(\frac{g_{k+1}}{\|g_{k+1}\|} - \lambda_k\frac{g_k}{\|g_{k+1}\|}\right)\right\|$$

is bounded away from zero. By (3.18) we have, therefore, the equality

$$|w_k'p_i| = O\left(\frac{|g_{k+1}'p_i|}{\|g_{k+1}\|} + \|g_k\|\,\frac{|g_k'p_i|}{\|g_{k+1}\|}\right)\ . \tag{3.20}$$

Combining (3.12), (3.13) and (3.20) we see that

$$|w_k'p_i| = O\left(\frac{\|g_k\|}{\|g_{k+1}\|}\left(\frac{|g_k'p_i|}{\|g_k\|} + \|v_{ik}\| + \|g_i\|\right)\right)\ . \tag{3.21}$$

Observing that by Lemma 3, $|\tau_k| = O(\|g_{k+1}\|/\|g_k\|)$ we deduce from (3.14), (3.15), (3.17), (3.19) and (3.21) the equality

$$|\tau_k d_i'q_k d_k'p_k| + |p_i'd_k| = O\left(\frac{|g_k'p_i|}{\|g_k\|} + \|v_{ik}\| + \|g_i\|\right)\ . \tag{3.22}$$

Since by Lemma 1, $\|s_{k+1}\|/g_k'p_k = O(\|g_{k+1}\|/\|g_k\|)$ the desired result follows now from (3.22) and Lemma 3 with $\ell = k+1$.

Using the above lemma we can now generalize a result obtained by Stoer [14] who proved that, for all update formulas considered in this paper, the sequence $\{\|x_j-z\|/\|x_{j-n}-z\|^2\}$ is bounded, provided x_0 is sufficiently close to z. Without requiring that x_0 is close to z we will prove the stronger result that the sequence $\{\|x_j-z\|/\|x_{j-n}-z\|^2\}$ converges to zero.

Theorem 2

Let $n \geq 2$. Then, for every update formula (2.4) with $\beta_1\beta_2 \geq 0$, $\beta_1 + \beta_2 \neq 0$,

$$\frac{\|g_j\|}{\|g_{j-n}\|^2} \to 0 \quad \text{as} \quad j \to \infty$$

$$\frac{\|x_j-z\|}{\|x_{j-n}-z\|^2} \to 0 \quad \text{as} \quad j \to \infty\ .$$

Proof:

Let $j \geq n$ and $j-n \leq i < k \leq j$. We will first show that

$$\frac{|g_k'p_i|}{\|g_k\|} = O\left(\frac{\|g_i\|^2}{\|g_k\|}\right) \tag{3.23}$$

and

$$\|v_{ik}\| = O\left(\frac{\|g_i\|^2}{\|g_{k-1}\|}\right)\ . \tag{3.24}$$

For $k = i+1$ the two statements follows from Lemma 1 and $v_{i,i+1} = 0$, respectively. Now suppose that $i+1 < k$ and (3.23) and (3.24) hold for some ν with $i < \nu < k$. By (3.12) and (3.13)

$$\begin{aligned}\frac{|g_{\nu+1}'p_i|}{\|g_{\nu+1}\|} &= O\left(\frac{\|g_\nu\|}{\|g_{\nu+1}\|}\left(\frac{|g_\nu'p_i|}{\|g_\nu\|} + \|v_{i\nu}\| + \|g_i\|\right)\right)\\ &= O\left(\frac{\|g_\nu\|}{\|g_{\nu+1}\|}\left(\frac{\|g_i\|^2}{\|g_\nu\|} + \frac{\|g_i\|^2}{\|g_{\nu-1}\|} + \|g_i\|\right)\right)\\ &= O\left(\frac{\|g_i\|^2}{\|g_{\nu+1}\|}\right)\ .\end{aligned}$$

Similarly by (3.9),

$$\|v_{i,\nu+1}\| = O\left(\frac{\|g_{\nu+1}\|}{\|g_\nu\|}\left(\frac{|g_\nu' p_i|}{\|g_\nu\|} + \|v_{i\nu}\|\right) + \|g_i\|\right)$$

$$= O\left(\frac{\|g_i\|^2}{\|g_\nu\|}\left(\frac{\|g_{\nu+1}\|}{\|g_\nu\|} + \frac{\|g_{\nu+1}\|}{\|g_{\nu-1}\|} + \frac{\|g_\nu\|}{\|g_i\|}\right)\right)$$

$$= O\left(\frac{\|g_i\|^2}{\|g_\nu\|}\right) .$$

This shows that (3.23) and (3.24) hold. Next we observe that

$$|p_k' G p_i| = |d_k' p_i + (p_k' G - d_k') p_i| \leq |d_k' p_i| + \|d_k - G p_k\| .$$

Using Lemma 1, (3.14), (3.23) and (3.24) we obtain, therefore, the relation

$$|p_k' G p_i| = O\left(\frac{\|g_i\|^2}{\|g_k\|}\right) . \tag{3.25}$$

Furthermore, it follows from (3.23) and (3.25) that, for $i = j-n,\ldots,k-1$ and $k = j-n+1,\ldots,j-1$,

$$|p_k' G p_i| = O\left(\frac{\|g_{j-n}\|^2}{\|g_{j-1}\|}\right) \quad \text{and} \quad \frac{|g_{j-1}' p_i|}{\|g_{j-1}\|} = O\left(\frac{\|g_{j-n}\|^2}{\|g_{j-1}\|}\right) . \tag{3.26}$$

To complete the proof we assume now that there are $\epsilon > 0$ and an infinite subset $J \subset \{0,1,\ldots,\}$ such that

$$\frac{\|g_j\|}{\|g_{j-n}\|^2} \geq \epsilon \quad \text{for} \quad j \in J . \tag{3.27}$$

Since $\|g_j\|/\|g_{j-1}\| \to 0$ as $j \to \infty$ this implies that

$$\frac{\|g_{j-n}\|^2}{\|g_{j-1}\|} \to 0 \quad \text{as} \quad j \to \infty,\ j \in J .$$

Hence it follows from (3.26) that there are a j_0 and a constant δ such that for $j \geq j_0$, $j \in J$ the matrix

$$P_j = (p_{j-1}, p_{j-2}, \ldots, p_{j-n})$$

is nonsingular and

$$\|P_j^{-1}\| \leq \delta . \tag{3.28}$$

Using Lemma 1, (3.12), (3.13), (3.24) and (3.26) we conclude that

$$|g_j' p_{j-1}| = O(\|g_{j-1}\|^2) = O\left(\frac{\|g_{j-1}\|^2}{\|g_{j-n}\|^2}\|g_{j-n}\|^2\right)$$

and for $i = j-2,\ldots,j-n$,

$$|g_j' p_i| = O\left(|1 - \sigma_{j-1}|\ \|g_{j-n}\|^2 + \frac{\|g_{j-n}\|^2}{\|g_{j-2}\|}\|g_{j-1}\| + \|g_{j-1}\|\ \|g_{j-n}\|\right) .$$

Since by Lemma 1, $|1 - \sigma_j| \to 0$ as $j \to \infty$ this implies that, for $i = j-1,\ldots,j-n$

$$\frac{g_j' p_i}{\|g_{j-n}\|^2} \to 0 \quad \text{as} \quad j \to \infty,\ j \in J .$$

In conjunction with (3.28) this shows that

$$\frac{\|g_j\|}{\|g_{j-n}\|^2} \to 0 \quad \text{as} \quad j \to \infty,\ j \in J .$$

Because this is a contradiction to (3.27) it follows that

$$\|g_j\|/\|g_{j-n}\|^2 \to 0 \quad \text{as} \quad j \to \infty ,$$

which in view of Lemma 1 completes the proof of the theorem.

As we have seen in Section 2 the matrix H_j can be represented in the form

$$H_j = \frac{p_j p_j'}{\rho_j g_j' p_j} + \sum_{i=2}^{n} \frac{p_{ij} p_{ij}'}{d_{ij}' p_{ij}} \tag{3.29}$$

where $\|p_j\| = \|p_{2j}\| = \ldots = \|p_{nj}\| = 1$. It is shown in the next lemma that an estimate for the rate of superlinear convergence can be obtain by using the numbers $\|d_{ij} - Gp_{ij}\|$, $i = 2, \ldots n$.

Lemma 5

Let H_j be given by (3.29). Then

i) $$\frac{|g'_{j+1}p_j|}{\|g_j\|} = O(\|g_j\|)$$

$$\frac{|g'_{j+1}p_{ij}|}{\|g_j\|} = O(\|d_{ij} - Gp_{ij}\| + \|g_j\|),\ i = 2,\ldots,n$$

ii) $$\frac{\|g_{j+1}\|}{\|g_j\|} = O(\max\left\{\frac{|g'_{j+1}p_j|}{\|g_j\|}, \frac{|g'_{j+1}p_{ij}|}{\|g_j\|}, i = 2,\ldots,n\right\}) .$$

Proof.

Replacing k with j and multiplying with p_{ij} we obtain from (3.10) the equality

$$\begin{aligned} g'_{j+1}p_{ij} &= g'_jp_{ij} - \sigma_jp'_{ij}Gs_j - \sigma_jp'_{ij}E_js_j \\ &= -\sigma_jd'_{ij}H_jg_j - \sigma_j(p'_{ij}G-d'_{ij})s_j - \sigma_jp'_{ij}E_js_j \\ &= -\sigma_j(p'_{ij}G-d'_{ij})s_j - \sigma_jp'_{ij}E_js_j \ , \end{aligned}$$

which by (3.11) implies

$$\frac{|g'_{j+1}p_{ij}|}{\|g_j\|} = O\left(\frac{\|\sigma_js_j\|}{\|g_j\|}(\|d_{ij}-Gp_{ij}\| + \|g_j\|)\right) = O(\|d_{ij}-Gp_{ij}\| + \|g_j\|) \ .$$

In connection with part iii) of Lemma 1 this completes the proof of the first part of the lemma.

Let $x \in E^n$ and

(3.30) $$x = \lambda\rho_jg_j + \sum_{i=2}^{n} \lambda_id_{ij} \ .$$

Since it follows from part v) of Assumption 1 that $\|H_j^{-1}\| \le 1/\tau_1$ we have

$$\|x\| = O(\max\{|\lambda|, |\lambda_2|,\ldots,|\lambda_n|\}) \ .$$

Multiplying both sides of (3.30) with p_j and p_{ij}, $i = 2,\ldots,n$, gives

$$p'_jx = \lambda\rho_jg'_jp_j \quad \text{and} \quad p'_{ij}x = \lambda_id'_{ij}p_{ij} \ .$$

Because $\rho_jg'_jp_j = p'_jH_j^{-1}p_j \ge 1/\tau_2$ and $d'_{ij}p_{ij} = p'_{ij}H_j^{-1}p_{ij} \ge 1/\tau_2$ this completes the proof of the lemma.

According to the above lemma we have

(3.31) $$\frac{\|g_{j+1}\|}{\|g_j\|} = O(\max\{\|d_{ij} - Gp_{ij}\|, i = 2,\ldots,n\} + \|g_j\|) \ .$$

It is interesting to observe that this relation is independent of the first term on the right hand side of (3.29), i.e., of $\|\rho_jg_j - Gp_j\|$.

It follows from (2.6) and part iv) of Lemma 1 that there is a representation of H_{j+1} in terms of n matrices of rank one containing the term

$$\frac{p_jp'_j}{d'_jp_j} \quad \text{with} \quad \|d_j - Gp_j\| = O(\|g_j\|) \ .$$

Similarly, an analoguous representation of H_{j+2} contains a term

$$\frac{p_{j+1}p'_{j+1}}{d'_{j+1}p_{j+1}} \quad \text{with} \quad \|d_{j+1} - Gp_{j+1}\| = O(\|g_{j+1}\|) \ .$$

This observation suggests that, under certain assumptions, it might be possible to prove that

(3.32) $$\max\{\|d_{ij} - Gp_{ij}\|, i = 2,\ldots,n\} = O(\|g_{j-n+1}\|)$$

which by (3.31) would imply

$$\frac{\|g_{j+1}\|}{\|g_j\|} = O(\|g_{j-n+1}\|) \quad . \tag{3.33}$$

Such a result can indeed be obtained if we assume that a certain lower bound on the rate of convergence, as specified in the following assumption, is valid.

Assumption 3

Let H_j be given by (3.29) and assume that there is $\delta > 0$ such that for every j

$$\frac{\|g_{j+1}\|}{\|g_j\|} \geq \delta \max_{i=2,\ldots,n} (\|d_{ij} - Gp_{ij}\|, \|g_{j-n+1}\|) \quad .$$

In view of (3.31) and the discussion leading to (3.32) Assumption 3 implies that the sequence $\{\|g_{j+1}\|/\|g_j\|\}$ does not converge faster than could be expected at best under the given circumstances for a general function $F(x)$.

As a first step towards establishing (3.33) we prove the following lemma.

Lemma 6

Let Assumption 3 be satisfied and let

$$H_j = \frac{p_j p_j'}{\rho_j g_j' p_j} + \sum_{i=2}^{n} \frac{p_{ij} p_{ij}'}{d_{ij}' p_{ij}} \quad .$$

Then there are a constant $\tau > 0$, independent of j, and vectors $d_{i,j+1}$, $p_{i,j+1}$, $i = 2,\ldots,n$, such that

i) $H_{j+1} d_{i,j+1} = p_{i,j+1}$, $\|p_{i,j+1}\| = 1$, $i = 2,\ldots,n$

ii) $H_{j+1} = \dfrac{p_{j+1} p_{j+1}'}{\rho_{j+1} g_{j+1}' p_{j+1}} + \displaystyle\sum_{i=2}^{n} \dfrac{p_{i,j+1} p_{i,j+1}'}{d_{i,j+1}' p_{i,j+1}}$

iii) $\|\rho_{j+1} g_{j+1} - Gp_{j+1}\| \leq \tau \max(\|g_j\|, \|g_{j+1}\|/\|g_j\|, \|d_{ij} - Gp_{ij}\|, i = 2,\ldots,n)$

iv) $\|d_{2,j+1} - Gp_{2,j+1}\| \leq \tau \|g_j\|$

v) $\|d_{i,j+1} - Gp_{i,j+1}\| \leq \tau \max(\|d_{i-1,j} - Gp_{i-1,j}\|, \|g_j\|)$, $i = 3,\ldots,n$.

Proof:

Let

$$w_j = \frac{g_{j+1} - \lambda g_j}{\|H_j(g_{j+1} - \lambda_j g_j)\|} , \quad q_j = \frac{H_j(g_{j+1} - \lambda_j g_j)}{\|H_j(g_{j+1} - \lambda_j g_j)\|}, \quad \lambda_j = \frac{g_{j+1}' p_j}{g_j' p_j} \quad .$$

Then $w_j' p_j = 0$ and

$$w_j \in \mathrm{span}(d_{2j},\ldots,d_{nj}) \quad .$$

By part v) of Assumption 1 this implies that

$$\|w_j - Gq_j\| = O(\max\|d_{ij} - Gp_{ij}\|, i = 2,\ldots,n)) \quad . \tag{3.34}$$

Define

$$\hat{p}_{1j} = \frac{p_j}{\sqrt{\rho_j g_j' p_j}} \quad , \quad \hat{p}_{ij} = \frac{p_{ij}}{\sqrt{d_{ij}' p_{ij}}} \quad , \ i = 2,\ldots,n$$

and set

$$P_j = (\hat{p}_{1j},\ldots,\hat{p}_{nj}) \quad , \quad v_j' = w_j' P_j \quad ,$$

$$z_j = \frac{\epsilon_j \|v_j\| e_n - v_j}{\|\epsilon_j \|v_j\| e_n - v_j\|}, \quad z_j' = (\zeta_1,\ldots,\zeta_n) \quad , \tag{3.35}$$

where

$$e_n' = (0,\ldots,0,1) \quad \text{and} \quad \epsilon_j = \begin{cases} 1 & \text{if} \quad (v_j)_n \leq 0 \\ -1 & \text{if} \quad (v_j) > 0 \quad . \end{cases}$$

Then

$$Q_j = I - 2 z_j z_j'$$

is a Householder - matrix with the property (see [10] for instance)

(3.36) $$Q_j v_j = \epsilon_j \|v_j\| e_n \ .$$

Let

(3.37) $$(\hat{q}_{1j},\ldots,\hat{q}_{nj}) = P_j Q_j = (\hat{p}_{1j} - 2\zeta_1 P_j z_j,\ldots,\hat{p}_{nj} - 2\zeta_n P_j z_j) \ .$$

Since $H_j = P_j P_j'$, we have

(3.38) $$P_j Q_j (P_j Q_j)' = P_j (Q_j Q_j') P_j' = P_j P_j' = H_j \ .$$

Furthermore, it follows from $w_j' p_j = 0$ and (3.36), respectively, that

(3.39) $$\hat{q}_{1j} = \hat{p}_{1j} \quad \text{and} \quad \hat{q}_{nj} = q_j \|\hat{q}_{nj}\| \ .$$

Defining

(3.40) $$p_{i,j+1} = \hat{q}_{i-1,j}/\|\hat{q}_{i-1,j}\| \ , \quad d_{i,j+1} = H_j^{-1} p_{i,j+1} \ , \quad i = 3,\ldots,n$$

we deduce from (3.37) through (3.40) that

$$H_j = \frac{p_j p_j'}{\rho_j g_j' p_j} + \frac{q_j q_j'}{w_j' q_j} + \sum_{i=2}^{n} \frac{p_{i,j+1} p_{i,j+1}'}{d_{i,j+1}' p_{i,j+1}} \ .$$

By (3.37) and (3.40) we have

(3.41) $$p_{i,j+1} = \frac{1}{\|\hat{q}_{i-1,j}\|} (\hat{p}_{i-1,j} - 2\zeta_{i-1} \sum_{\nu=1}^{n} \zeta_\nu \hat{p}_{\nu j}), \ i = 3,\ldots,n$$

(3.42) $$d_{i,j+1} = \frac{1}{\|\hat{q}_{i-1,j}\|} (H_j^{-1}\hat{p}_{i-1,j} - 2\zeta_{i-1} \sum_{\nu=1}^{n} \zeta_\nu H_j^{-1}\hat{p}_{\nu j}), \ i = 3,\ldots,n \ .$$

Since $\zeta_1 = 0$, $|\zeta_\nu| \le 1$, $\nu = 2,\ldots,n$, $\hat{q}_{ij}' H_j^{-1} \hat{q}_{ij} = 1$ and

$$\|H_j^{-1}\hat{p}_{ij} - G\hat{p}_{ij}\| = O(\|d_{ij} - Gp_{ij}\|), \ i = 2,\ldots,n \ ,$$

it follows from (3.41) and (3.42) that, for $i = 3,\ldots,n$,

(3.43) $$\|d_{i,j+1} - Gp_{i,j+1}\| = O(\|d_{i-1,j} - Gp_{i-1,j}\| + |\zeta_{i-1}| \sum_{\nu=2}^{n} \|d_{\nu j} - Gp_{\nu j}\|) \ .$$

Observing that

$$\|\epsilon_j \|v_j\| e_n - v_j\|^2 \ge \|v_j\|^2 = w_j' P_j P_j' w_j = w_j' q_j$$

is bounded away from zero we deduce from (3.35) the equality

(3.44) $$|\zeta_i| = O(|(v_j)_i|) = O(|w_j' p_{ij}|), \ i = 2,\ldots,n-1 \ .$$

Since by assumption $\|g_{j+1}\| \ge \delta \|g_j\| \ \|g_{j-n+1}\|$ it follows from Lemma 1 that

$$\lambda_j = g_{j+1}' p_j / g_j' p_j \to 0 \quad \text{as} \quad j \to \infty$$

which as in the proof of Lemma 4 implies that

$$\left\| H_j \left(\frac{g_{j+1}}{\|g_{j+1}\|} - \lambda_j \frac{g_j}{\|g_{j+1}\|} \right) \right\|$$

is bounded away from zero. Thus we conclude from (3.44), $g_j' p_{ij} = 0$, and Lemma 5 that, for $i = 2,\ldots,n-1$,

(3.45) $$|\zeta_i| = O\left(\frac{|g_{j+1}' p_{ij}|}{\|g_{j+1}\|} \right) = O\left(\frac{\|d_{ij} - Gp_{ij}\| + \|g_i\|}{\max\{\|d_{ij} - Gp_{ij}\|, i=2,\ldots,n} \right) \ .$$

Therefore, part v) of the lemma follows from (3.43) and (3.45).

By (2.6)

$$H_{j+1} = \frac{p_j p_j'}{d_j' p_j} + \omega_j \frac{u_j u_j'}{w_j' u_j} + \sum_{i=3}^{n} \frac{p_{i,j+1} p_{i,j+1}'}{d_{i,j+1}' p_{i,j+1}} \ ,$$

where

$$u_j = \frac{q_j + \alpha_j p_j}{\|q_j + \alpha_j p_j\|}, \ \alpha_j = -\frac{d_j' q_j}{d_j' p_j} \ .$$

Since by Lemma 1 and (2.7), $|\alpha_j| = O(\|g_{j+1}\|/\|g_j\|)$, $|\omega_j - 1| = O(\|g_{j+1}\|/\|g_j\|$

it follows that

(3.46) $$\left\|\frac{w_j}{\omega_j} - Gu_j\right\| = O(\|w_j - Gq_j\| + \|g_{j+1}\|/\|g_j\|) \quad .$$

Furthermore, $g_{j+1} \in \operatorname{span}\{d_j, w_j\}$ implies that

(3.47) $$\rho_{j+1} g_{j+1} = \lambda_j d_j + t_j \frac{w_j}{\omega_j}, \; p_{j+1} = \lambda_j p_j + t_j u_j \quad .$$

Observing that by Lemma 1, $\|d_j - Gp_j\| = O(\|g_j\|)$ and $|\lambda_j| \le \bar{\lambda}$, $|t_j| \le \bar{t}$ for some $\bar{\lambda}$ and $\bar{t}$, independent of j, we obtain from (3.46) and (3.47) the relation

(3.48) $$\|\rho_{j+1} g_{j+1} - Gp_{j+1}\| = O(\|w_j - Gq_j\| + \|g_j\| + \|g_{j+1}\|/\|g_j\|) \quad .$$

Finally, let

(3.49) $$p_{2,j+1} = \bar{\lambda}_j p_j + \bar{t}_j p_{j+1}$$

be such that $\|p_{1,j+1}\| = 1$ and $g'_{j+1} p_{2,j+1} = 0$. Then

(3.50) $$|\bar{t}_j| = |\bar{\lambda}_j| \frac{|g'_{j+1} p_j|}{g'_{j+1} p_{j+1}} = O\left(|\bar{\lambda}_j| \frac{\|g_{j+1}\|}{g'_{j+1} p_{j+1}} \frac{\|g_j\|^2}{\|g_{j+1}\|}\right) = O\left(\frac{\|g_j\|^2}{\|g_{j+1}\|}\right) \quad ,$$

where the last two equalities follows from Lemma 1 and the fact that $\{\bar{\lambda}_j\}$ is bounded. Therefore, defining

$$d_{2,j+1} = H^{-1}_{j+1} p_{2,j+1}$$

we obtain from (3.48), (3.49) and (3.50)

(3.51) $$\|d_{2,j+1} - Gp_{2,j+1}\| \le |\bar{\lambda}_j| \, \|d_j - Gp_j\| + |\bar{t}_j| \, \|\rho_{j+1} g_{j+1} - Gp_{j+1}\|$$
$$= O(\|g_j\| + |\bar{t}_j|(\|w_j - Gq_j\| + \|g_j\| + \|g_{j+1}\|/\|g_j\|))$$
$$= O\left(\|g_j\| + \frac{\|g_j\|}{\delta v_j}(\|w_j - Gq_j\| + \|g_j\|)\right)$$
$$= O(\|g_j\|) \quad ,$$

where

$$v_j = \max\{\|g_{j-n+1}\|, \|d_{ij} - Gp_{ij}\|, \; i = 2,\ldots,n\}$$

and the last equality follows from (3.34) and Assumption 3.

Since $g'_{j+1} p_{2,j+1} = 0$ we can now represent H_{j+1} in the form

$$H_{j+1} = \frac{p_{j+1} p'_{j+1}}{\rho_{j+1} g'_{j+1} p_{j+1}} + \sum_{i=2}^{n} \frac{p_{i,j+1} p'_{i,j+1}}{d'_{i,j+1} p_{i,j+1}} \quad .$$

In conjunction with (3.48), and (3.51) this completes the proof of the lemma.

A repeated application of Lemma 6 shows that the estimate (3.32) is valid and leads to the following theorem.

Theorem 3

Let Assumption 1 and 3 be satisfied. Then for every update formula (2.4) with $\beta_1 \beta_2 \ge 0$, $\beta_1 + \beta_2 \ne 0$ the following statements hold.

i) $$\frac{\|x_{j+1} - z\|}{\|x_j - z\|} = O(\|x_{j-n+1} - z\|)$$

ii) $$\|H_j - G^{-1}\| = O(\|x_{j-n} - z\|)$$

iii) $$|1 - \sigma_j| = O(\|x_{j-n} - z\|) \quad .$$

Proof:

If we write each H_j in the form

$$H_j = \frac{p_j p_j'}{\rho_j g_j' p_j} + \sum_{i=2}^{n} \frac{p_{ij} p_{ij}'}{d_{ij}' p_{ij}}$$

it follows from Lemma 6 that

(3.52) $$\|d_{ij} - Gp_{ij}\| = O(\|g_{j-i+1}\|),\ i = 2,\dots,n \quad ,$$

which by Lemma 5 implies

(3.53) $$\frac{\|g_{j+1}\|}{\|g_j\|} = O(\|g_{j-n+1}\|) \quad .$$

The first statement of the theorem follows now from (3.53) and part v) of Lemma 1. Furthermore, we obtain from (3.52), (3.53) and part iii) of Lemma 6 the relation

(3.54) $$\|\rho_j g_j - Gp_j\| = O(\|g_{j-n}\|) \quad .$$

Observing that, by Lemma 1, the sequence

$$\{(p_j, p_{1j}, \dots, p_{nj})^{-1}\},\ j = 1,2,\dots$$

exists and is bounded we deduce from (3.52) and (3.54) that

$$\|H_j - G^{-1}\| = O(\|g_{j-n}\|) = O(\|x_{j-n} - z\|) \quad .$$

Finally, it follows from Taylor's theorem that there is

$$y_j \in \{x \mid x = x_j - ts_j,\ 0 \le t \le 1\}$$

such that

$$2[F(x_j - s_j) - F(x_j) + g_j s_j] = s_j' G(y_j) s_j \quad .$$

Because

$$s_j' G(y_j) s_j = g_j' s_j + s_j' G(H_j - G^{-1}) g_j + s_j'(G(y_j) - G) s_j$$

we obtain from the definition of σ_j (see (2.9)) the relation

$$|1 - \sigma_j| = \frac{|s_j' G(y_j) s_j - g_j' s_j|}{s_j' G(y_j) s_j}$$

$$= O\left(\frac{\|g_j\|}{\|s_j\|} \|G\|\ \|H_j - G^{-1}\| + \|G(y_j) - G\|\right)$$

$$= O(\|x_{j-n} - z\|) \quad ,$$

where the last equality follows from

$$\|G(y_j) - G\| \le L \max(\|x_j - z\|,\ \|x_j - s_j - z\|) = O(\|x_j - z\|) \quad .$$

As a further consequence of Assumption 3 we have the following theorem which implies that n consecutive search directions are asymptotically conjugate with respect to the Hessian matrix of $F(x)$ at z.

Theorem 4

Let Assumptions 1 and 3 be satisfied. Then for every update formula (2.4) with $\beta_1 \beta_2 \ge 0$, $\beta_1 + \beta_2 \ne 0$ the following statements hold.

i) $\|H_j d_i - p_i\| = O(\|g_i\|)$, $i = j-1,\dots,j-n$

ii) $\dfrac{|g_k' p_i|}{\|g_k\|} = O\left(\dfrac{\|g_i\|}{\|g_{k-n}\|}\right)$, $j-n \le i < k \le j$

iii) $|p_k' G p_i| = O\left(\dfrac{\|g_i\|}{\|g_{k-n}\|}\right)$, $j-n \le i < k < j$.

Proof:

Let $j \ge n$ and $j-n \le i < k \le j$. We will first show that

$$\frac{|g_k'p_i|}{\|g_k\|} = O\left(\frac{\|g_i\|}{\|g_{k-n}\|}\right) \text{ and } \|v_{ik}\| = O(\|g_i\|) \quad . \tag{3.55}$$

For $k = i+1$ the two statements follow from Lemma 1 and Lemma 3, respectively. Suppose that $i+1 < k$ and (3.55) holds for some $i < \nu < k$. By (3.12), (3.13), and Theorem 3

$$\frac{|g_{\nu+1}'p_i|}{\|g_{\nu+1}\|} = O\left(|1-\sigma_\nu| \frac{\|g_\nu\|}{\|g_{\nu+1}\|} \frac{|g_\nu'p_i|}{\|g_\nu\|} + \frac{\|g_\nu\|}{\|g_{\nu+1}\|}\left(\|v_{i\nu}\| + \|g_i\|\right)\right)$$

$$= O\left(\frac{\|g_{\nu-n}\|}{\|g_{\nu-n+1}\|}\frac{\|g_i\|}{\|g_{\nu-n}\|} + \frac{\|g_i\|}{\|g_{\nu-n+1}\|}\right)$$

$$= O\left(\frac{\|g_i\|}{\|g_{\nu-n+1}\|}\right) \quad .$$

Similarly by (3.9) and Theorem 3

$$\|v_{i,\nu+1}\| = O\left(\frac{\|g_{\nu+1}\|}{\|g_\nu\|}\left(\frac{|g_\nu'p_i|}{\|g_\nu\|} + \|v_{i\nu}\|\right) + \|g_i\|\right)$$

$$= O\left(\|g_{\nu-n+1}\|\left(\frac{\|g_i\|}{\|g_{\nu-n}\|} + \|g_i\|\right) + \|g_i\|\right)$$

$$= O(\|g_i\|) \quad .$$

Since $v_{ij} = H_j d_i - p_i$ this completes the proof of the first two parts of the theorem. Finally we observe that

$$|p_k'Gp_i| = |d_k'p_i + (p_k'G-d_k')p_i| \le |d_k'p_i| + \|d_k - Gp_k\| \quad ,$$

which by (3.14), Lemma 1 and the first two parts of the theorem implies

$$|p_k'Gp_i| = O\left(\frac{|g_k'p_i|}{\|g_k\|} + \|v_{ik}\| + \|g_i\|\right) = O\left(\frac{\|g_i\|}{\|g_{k-n}\|}\right) \quad .$$

REFERENCES

[1] C. G. Broyden, "Quasi-Newton methods and their application to function minimization", Mathematics of Computation 21 (1967) pp. 368-381.

[2] C. G. Broyden, "The convergence of a class of double-rank minimization algorithms", Parts 1 and 2, Journal of the Institute of Mathematics and its Applications 6 (1970) pp. 76-90, pp. 222-231.

[3] W. Burmeister, "Die Konvergenzordnung des Fletcher-Powell-Algorithmus", Zeitschrift für Angewandte Mathematik and Mechanik 53 (1973) pp. 693-699.

[4] W. C. Davidon, "Variable metric methods for minimization", Argonne National Laboratories rept. ANL-5990 (1959).

[5] L. C. W. Dixon, "Variable metric algorithms: necessary and sufficient conditions for identical behaviour on non-quadratic functions", Journal of Optimization Theory and Applications 10 (1972) pp. 34-40.

[6] R. Fletcher and M. J. D. Powell, "A rapidly convergent descent method for minimization", The Computer Journal 6 (1963) pp. 163-168.

[7] R. Fletcher, "A new approach to variable metric algorithms", The Computer Journal 13 (1970) pp. 317-322.

[8] D. Goldfarb, "A family of variable metric methods derived by variational means", Mathematics of Computation 24 (1970) pp. 23-26.

[9] H. Y. Huang, "Unified approach to quadratically convergent algorithms for function minimization", Journal of Optimization Theory and Applications 5 (1970) pp. 405-423.

[10] B. Noble, "Applied linear algebra", Prentice-Hall, 1969.

[11] K. Ritter, "Global and superlinear convergence of a class of variable metri methods", Mathematics Research Center TSR #1945, University of Wisconsin-Madison 1979.

[12] D. F. Shanno, "Conditioning of quasi-Newton methods for function minimization", Mathematics of Computation 24 (1970) pp. 647-656.

[13] G. Schuller, "On the order of convergence of certain quasi-Newton Methods", Numerische Mathematik 23 (1974) pp. 181-192.

[14] J. Stoer, "On the convergence rate of imperfect minimization algorithms in Broyden's β-class", Mathematical Programming 9 (1975) pp. 313-335.

KR/jvs

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
#1950

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
6 On the Rate of Superlinear Convergence of a Class of Variable Metric Methods

5. TYPE OF REPORT & PERIOD COVERED
Summary Report - no specific reporting period

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
10 Klaus Ritter

8. CONTRACT OR GRANT NUMBER(s)
15 DAAG29-75-C-0024

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Mathematics Research Center, University of Wisconsin
610 Walnut Street
Madison, Wisconsin 53706

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Work Unit Number 5 - Mathematical Programming and Operations Research

11. CONTROLLING OFFICE NAME AND ADDRESS
U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, North Carolina 27709

12. REPORT DATE
11 Apr[redacted]1979

13. NUMBER OF PAGES
33

12 32p

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
14 MRC-TSR-1950

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

9 Technical summary rept.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Unconstrained minimization, variable metric method, superlinear convergence.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper considers a class of variable metric methods for unconstrained minimization. Without requiring exact line searches each algorithm in this class converges globally and superlinearly. Various results on the rate of the superlinear convergence are obtained.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

221 200

JOB